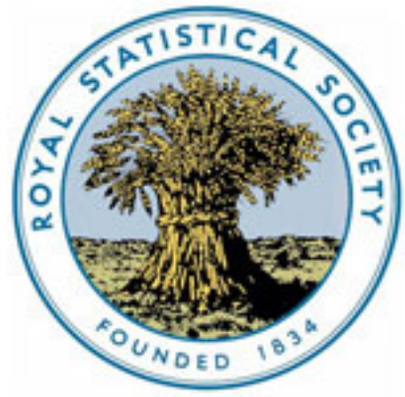



*I.S./I.C.I. Conference*

# *Statistical Methods in Research and Production*

G. A. COUTIE

## 1. Introduction

The title of this talk is the same as that of one of the I.C.I. books on statistics, with which I am sure you will be familiar. However, the content will be quite different, and I shall try to describe the extent to which established statistical techniques are applied to the diverse technical activities of I.C.I., and also to indicate those areas where more novel methods are required. I should point out that the word "research" is being used here to cover all technical activities that are not directly related to production processes, and that it covers the important fields of technical service and laboratory process development.

The role of the statistician on the technical side is to collaborate with the chemist (or physicist or engineer as the case may be) in the design of experimental work and in the analysis of data. Many of the more fundamental statistical ideas are of universal use in any quantitative study, and the chemist is encouraged to use such methods on his own problems, thereby avoiding the time-consuming task of explaining the details of his work to the statistician. Indeed it must be regarded as part of the job of the professional statistician to ensure that the people with whom he works appreciate the implications of the word variability, and it is equally important that they should be kept informed of any new statistical techniques that may be useful to them. Most Divisional Statistical Sections in I.C.I. therefore undertake a certain amount of teaching, and at General Chemicals Division, for example, a series of 4-day statistical courses has recently been given to a total of about 100 chemists. In these courses great emphasis has been laid on practical work where descriptions have been given of such techniques as cumulative sum charts, evolutionary operation and the use of a computer for multiple regression, and a "black box" approach has been used so that chemists can compare different methods for the sequential planning of experiments. Because of the current shortage of adequately trained statisticians it is essential that they should be employed on the more complex problems and not to provide a routine service on straightforward applications.

You will be well aware that the technical interests of I.C.I. are not restricted to the chemical field, and in discussing the application of

statistical methods in our research activities I will split these into the following four categories—

(i) chemical
(ii) biological
(iii) agricultural
(iv) physical and engineering.

Although many of the simpler techniques are common to all four, each has its own particular requirements from a statistical point of view.

## 2. Chemical Research

Many of the manufacturing processes in the company take the form of a chemical reaction to produce a crude product, which then undergoes one or more physical processes such as distillation or filtration, and it is natural that a large part of the Company's research activities is associated with the derivation and development of chemical reactions. This work is normally the responsibility of the Research Department in each Division, but several Divisions have Experimental or Process Investigation Departments attached to the Works who carry out process development work.

In the inventive stage of chemical research, when completely new territories are being explored, it is not easy for statistical methods to be of any use other than to assist in screening procedures. Once the chemical route to a product has been decided upon, however, the most frequently occurring problem is that of choosing the levels of the controllable factors such as time, temperature, and concentration in order that some response such as the yield or cost should be optimized, and in planning such work much use is made of factorial experimentation and of the response surface designs developed by G. E. P. Box. In fact Box's work on response surfaces was carried out specially to meet the needs of chemical research workers which he appreciated while working for our Dyestuffs Division where, because of the nature of the manufacturing activities, several hundred chemical reactions are studied annually. As with many other theories the exploitation of response surface designs leads to practical difficulties. One of these is the choice of the variables and the relative scales on which they are measured, where the chemist is asked by the statistician to arrange that the least squares response surface should be well-conditioned and approximately a second degree surface. The success or otherwise of chemical experimentation often depends as much as anything on the skill of the chemist in defining the right region of experimentation. Another difficulty arises from the fact that a research chemist cannot always achieve the experimental conditions that the statistician would like him to, which leads to a

non-orthogonal design. Now that electronic computers instead of desk calculating machines are used for response surface calculations it can safely be admitted that the only virtue of exact orthogonality is ease of calculation, which the computer does not need.

Perhaps the most serious criticism of factorially based experiments is the fact that they are essentially fixed sample size designs, involving a decision before the work starts on the particular sets of conditions to be tried. They are ideal for those situations where the results all come to hand at the same time, as for example in the agricultural experiments for which they were initially evolved, but the research chemist quite rightly thinks of his experimentation as being sequential and rebels against anyone who asks him to complete a "box of eight" trials before any statistical analysis is attempted. It has always been difficult to interest the research chemist in experimental design because of this, and a working team on the sequential planning of experiments has recently been set up by the Statistical Methods Panel to explore methods whereby the chemist, the statistician and the computer may collaborate more successfully in chemical experimentation.

The vast majority of chemical investigations are necessarily empirical, but some of the company's statisticians are currently working on the difficult problem of the design and analysis of experiments to determine the kinetic rate constants in a chemical reaction. This involves problems of non-linear estimation, in which the use of an electronic computer is essential.

### 3. Biological Research

The majority of biological experimentation is handled by Pharmaceuticals Division in their new laboratories at Alderley Park, and many of the problems in this type of work are quite distinct from those in all other categories.

The effectiveness of a compound against a particular disease has usually to be assessed in animals infected with the disease, and when the response is one of two possibilities, e.g. cure or no cure, death or survival, use is made of the techniques of probit analysis or logit analysis. These techniques are used particularly for estimating and comparing the toxicity of drugs which may be expressed by the dose which kills 50% of animals (L.D.50), the dose which cures 50% of animals (E.D.50) and the therapeutic ratio (L.D.50/E.D.50), giving a measure of the degree of safety in the use of the drugs. Whenever the response is quantitative, e.g. the extent of the cure, use is made of regression and covariance techniques.

Another specialized problem in this field is the design of biological screening procedures, which occur when a large number of chemical compounds are to be tested on animals in the laboratory in order to

find those which are suitable for use as drugs for the treatment of diseases. Several thousand compounds are examined each year at Alderley Park, for up to 30 desirable properties. When deciding on the number of animals or the amount of testing capacity to use on each compound a compromise must be made between two opposing requirements. On the one hand more testing per compound will increase the chance of spotting an active compound, but on the other hand it will reduce the number of compounds that can be tested. Several statistical criteria are available to ensure that a reasonable compromise is reached, and these are used as an integral part of the research activities of our Pharmaceuticals Division.

## 4. Agricultural Research

I.C.I.'s research on agricultural problems is centred on Jealott's Hill, which houses the Agricultural Research & Development Department of Billingham Division and also the research and development laboratories of Plant Protection Limited. One statistical unit deals with the experimental work of all those departments, and another handles agricultural econometric questions for Billingham Division.

Most of the pioneer work on experimental design was carried out at Rothamsted in connection with agricultural research, so the work at Jealott's Hill falls largely into established patterns. Typical examples are field experiments with fertilizers on a variety of crops, or trials to compare different systems of manuring or of cultivation, carried out either at Jealott's Hill itself or, by I.C.I. staff, on private farms throughout Britain. The design and interpretation of such experiments, and of similar ones on a laboratory scale, form a large part of the statistical work.

Research on chemicals for the control of weeds, pests, and plant diseases provides problems which combine some aspects of traditional agricultural statistical methods with some of those described in the previous section. Similar problems of biological screening can arise, and analogous methods may be used in designing standard techniques for testing and comparing different formulations of biologically active materials. Experimental work on a field scale often presents more statistical difficulties where pests and diseases, rather than the crop itself, are concerned, since lack of uniformity and of satisfactory controls has often to be overcome.

Also at Jealott's Hill Research Station is a small unit concerned with the statistical analysis of about 240 farm cost accounts and other farming records obtained mainly from grassland farms, some primarily producing milk and some producing meat. Most of the farms are in England and Wales and are predominantly in the wetter western parts of the country. The data are obtained through the co-operation of commercial farmers who are chosen because they are

following, or wish to follow, a policy of intensifying their use of grassland, and in this sense the group of farms form a purposive sample. The main object of the analyses is to help determine the technique necessary to use farming resources most profitably and to assess the input-output functions associated with the use of these techniques. Naturally the statistical methods used fall largely into the field of regression analysis, but solutions of input-output problems have also been sought through systems of simultaneous equations by the limited information, single equation technique (L.I.S.E.) using the Elliott 402 Computer at Billingham. Two papers concerned with this work have, in fact, been published in your Journal.

## 5. Physical and Engineering Research

In the physical and engineering fields there is a wide range of work that is amenable to statistical treatment. This is in fact a miscellaneous category, grouping together all the problems that arise in any well-organized laboratory in which variability is accepted as something to be lived with rather than buried. It includes the universal problems of finding out how accurate are the instruments that we measure with, and how reliable are the materials we use, as well as topics which are more particular to our own Company concerned with the physical testing of the products that we make, and the design and reliability of the plants in which we make them. Designed experiments play an important part in these activities, but many other specialized techniques such as sequential testing, extreme value theory and subjective analyses find their way into particular problems. I would venture the guess, however, that in several Divisions of I.C.I. it is the enormous variety in the nature of the problems, just as much as the statistical techniques used, which keeps the statistician interested in—perhaps fascinated by—his job.

A few examples from recent work should suffice to indicate the range of topics that the I.C.I. statistician may meet. In fertilizer manufacture many thousands of paper sacks are filled each day, and regression methods and discriminant functions have been used to identify factors associated with high burstage rates. The value of high-styrene resins for reinforcing shoe soling was investigated by measuring the physical properties of twelve resin-rubber compounds after a statistically planned wearing trial in which 78 schoolboys participated, and multiple comparison methods were used to assess differences between the resins, and their advantage over leather.

In order to determine whether the extra expense of a 4-line pneumatic control system was justified in comparison with a 2-line system a fractional factorial experiment was carried out in which the effects of 4 control variables were measured for each of 4 different lengths of air line. Colour matching plays an integral part in the activities of

both our Paints and Dyestuffs Divisions, and statistical methods are used to relate the performance of instruments measuring colour on a 3-co-ordinate system to the more subjective better tried method of using the human eye. In designing a multi-product plant a Monte Carlo computer simulation was used in order to estimate the likely amount of drying capacity needed under a variety of production programmes.

These are only a few examples drawn from throughout the Company. In each Division there is naturally a bias towards one or two particular types of application, but the existence of the Statistical Methods Panel and its various study groups means that there are easy opportunities for an interchange of advice and views between different Divisions.

## 6. Statistical Methods in Production

On the production side of the Company's activities statistical methods play an important part in the control of manufacturing processes. The extent to which a process can be controlled, and the way in which the control is achieved, must depend to a great extent on the nature of the process, and the result is that widely differing methods and organizations are used. The largest quality control organization in I.C.I. is in Fibres Division in connection with the manufacture of "Terylene" at our Wilton Works. This organization is directed by the Division Quality Manager and includes two statisticians amongst its graduate staff; altogether it constitutes some 10% of the total works strength. Its activities, however, cover a very wide field, including laboratories for raw materials, intermediates and fibres testing, product inspection and packing, process standardization checks, complaint analysis, and textile monitoring as well as the operation of numerous statistically based procedures for process control and acceptance/rejection purposes. In other Divisions, where the manufacture covers a wide range of products, the responsibility for quality control is not so centralized as it is in Fibres Division, and the Divisional statisticians collaborate with individual works or plant managers in the setting up of suitable control schemes. An important field of application in some Divisions is in the control of the accuracy of weighing equipment, and in many analytical laboratories statistical techniques are used in the control of the standard and the reproducibility of routine analysis.

As far as techniques are concerned, increasing use is being made of control schemes based on cumulative sums. A lot of work has been carried out by some of the company statisticians, helped by Professor G. A. Barnard, into the theory and practice of this type of control, which was suggested in the first place by Dr. E. S. Page. Our conclusions are that a cumulative sum control scheme will almost

always be preferable to the more conventional Shewhart chart both on the grounds of efficiency in detecting changes in the mean level of a response, and in ease of operation. Both the numerical type of scheme suggested by Ewan and Kemp, from British Nylon Spinners, and the graphical type suggested by Professor Barnard, are in widespread use. With chemical processes it is sometimes not easy to correct or control the process when the output falls away from specification because the relationships between input and output are not known. In these circumstances it is still useful to keep a chart running to show what is happening to the output, merely for information purposes, and the cumulative sum chart quickly proves to be popular with production people for this purpose because of the startling way in which it demonstrates a change in mean.

On a full production scale it is not always practicable to carry out planned experimentation, and in order to relate the input and output of a plant use must be made of regression procedures. It is well known among statisticians that these are no substitute for deliberately-made variations, and multiple regression analysis has borne a social stigma for many years. With the aid of fast, comprehensive computer programmes for multiple regression analysis, however, which will accept many input variables, process transformations of these variables and go through a procedure for eliminating those which do not materially affect the output variable, it is tempting to try regression analysis first on many problems in order that the maximum amount of information may be extracted from the masses of routine plant data that are usually available. The result is that the programme for multiple regression on the "Mercury" Computer is well used, although it is recognized that when no deliberate variations have been made the chances of gaining really useful information are slender. The major effort involved in this work is the punching of plant records on to paper tape prior to processing on the computer. Much thought is being given to the possibility of carrying out routine summaries of process records on a computer, and in fact one Division has already installed an Elliott 803 Computer for this purpose. Where records are automatically punched on to tape the use of regression procedures is very much simplified.

As part of the effort that is continually being made to improve the performance of a production unit, the method of evolutionary operation is sometimes used. This is a technique of process operation that was developed by G. E. P. Box while he was with I.C.I., involving continual planned variations in the process conditions. In recent years it has received a considerable amount of publicity both in this country and, more especially, in the U.S.A. It is perhaps unfortunate that the tone of much that has been said on this subject has tended to give the impression that it is a technique which may be applied

successfully to almost all chemical plants, whereas most of the published examples of applications are of the same basic type in which (for example) the temperature of a reaction and the concentrations of one of the reactants are varied at two levels each in repeated cycles round a $2 \times 2$ factorial design. The experience of I.C.I. production chemists who have tried to make use of evolutionary operation is that the practical difficulties of being able to measure the required input and output variables satisfactorily, and of ensuring that with campaign-wise manufacture the programme carries on long enough to make it worthwhile, often far exceed the difficulties of designing and analysing the work. As a result the successful applications in I.C.I. are limited to about 40, out of a total of several thousand chemical processes. About two-thirds of the applications are in Dyestuffs Division where, because of the relatively small scale on which some of the products are manufactured, the value of a 1% increase in yield may be only about £1,000 p.a. We cannot claim, therefore, that evolutionary operation has been an outstanding success so far, and it is unlikely that the savings in the Company that can be directly attributed to it exceed £50,000 p.a.

Despite this limited number of applications the technique of evolutionary operation is undoubtedly useful in those situations where it can fit in easily with production requirements, and as improvements are made to the instrumentation of chemical plants, and in methods of chemical analysis, the scope for gaining information from continual planned variations in the process will clearly widen. An alternative method for designing the variations has been suggested and tried by Mr. Spendley and Dr. Himsworth of our Billingham Division. They refer to it as simplex evolutionary operation, and it is particularly suited to those situations in which the rate of flow of information from the plant is so high that an automatic method of planning and analysing variations is called for. One of the difficulties we are conscious of is that of who should do the work that is associated with evolutionary operation. The production chemist who has received little or no statistical training claims that he has enough to do in running his plant and cannot handle the method as a routine. The statistician does not want to be involved with it full-time because the technique is not particularly interesting from a statistical point of view. In America chemical companies have made much more use of evolutionary operation than we have because they have far more chemists and chemical engineers who have received a fair amount of training in the application, if not the theory, of statistical methods. It is significant that the greatest use of evolutionary operation is in our factory at Huddersfield, where a chemist with statistical training is employed full-time in designing and running applications of this technique throughout the works.

No discussion of the use of statistical methods in production problems can finish without a mention of the use of computers for simulation procedures. This is a rapidly expanding field, and a good deal of experience has been built up over the last few years by several Divisions of I.C.I. The list of operations that have been simulated grows all the time, and ranges from individual chemical processes to virtually a whole factory. Perhaps the inclusion of this as a field of statistical interest is stretching things a bit, since the only statistical concept in many applications is that of sampling randomly from an observed distribution, and in other companies simulation techniques would come under the heading of operations research. However, in I.C.I., for better or for worse, we do not have any central operations research organization, and in only one Division is there a department with this title. Such work is often carried out by statisticians, who must necessarily work closely with computer experts. In fact it is found that an increasing number of the problems that are submitted to statisticians in I.C.I. necessitate the use of many techniques other than mathematical statistics, and in particular that the statistician must regard a knowledge of electronic computers as part of his equipment.

In the next talk Dr. Baines will discuss in more detail the impact that computers have had on statistics.